Exhibit B

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA

IN RE: UBER TECHNOLOGIES, INC., PASSENGER SEXUAL ASSAULT LITIGATION	MDL No. 3084 CRB

DECLARATION OF JONATHAN JAFFE

I, Jonathan Jaffe, declare:

1. I am a technology consultant, data scientist, and the founder and owner of Its-Your-Internet, an advanced technology software, data, and litigation support consulting firm I established in 2008.

2. I provide ESI[1] discovery and litigation consulting for attorneys in complex litigation throughout the litigation lifecycle from inception to ultimate resolution.

3. In multi-district litigations ("MDLs") and actions inclusive of this one, I have assisted with discovery efforts regarding the structure of corporate ESI and methods by which to produce voluminous data. These corporate systems included the primary systems for multiple Fortune 500 companies including for the three major credit bureaus, the largest pharmaceutical companies, large banks, and technology companies such as Google.

4. I have spent 30 years, dating back to 1993, building and evaluating large-scale distributed systems, both consulting directly for corporations and in the context of litigation. This includes extensive experience with data analysis, working with both structured and unstructured data.[2]

---

[1] "ESI" or Electronically Stored Information encompasses both documents and data

[2] Structured data refers to data that is organized and formatted in a predefined and standardized manner, typically in a database or spreadsheet. Structured data is usually composed of discrete data fields that are defined by their data type, such as numbers, dates, or text. Structured data can be easily searched, analyzed, and processed using various software

5. In roles both as a consulting and testifying expert, I have worked to filter, join, and summarize government and corporate datasets that have had *hundreds of millions* to **hundreds of billions** of records, including in the Opioid Litigations,[3] various Medical Device and Drug Litigations,[4] various Environmental and Toxic tort litigations, and Class Action cases involving alleged violations of the Fair Credit Reporting Act ("FCRA"), the Electronic Funds Transfer Act ("EFTA"), the Telephone Consumer Protection Act ("TCPA"), and the Fair Labor Standards Act ("FLSA"), unjust enrichment and contractual interference. I have been asked to examine voluminous data in multiple federal class action cases in which I assisted in identifying class members.

6. Between 2003 and 2008, I worked in-house for Weitz & Luxenberg, P.C. overseeing software development, the submission of hundreds of thousands of claims in mass tort litigation and testifying as an expert witness in federal cases.

7. I graduated magna cum laude from Columbia University with a BA in Economics-Mathematics in 1999.

8. A copy of my *curriculum vitae* is attached hereto as **Exhibit A**.

9. I have personal knowledge of the statements set forth below and could testify as to same if called upon to do so.

---

tools and techniques, and is commonly used for transactional data. Semi-structured or unstructured data refers to data that is not organized or formatted in a predefined manner, and may include text, images, audio, video, or other forms of data that are not typically analyzed using traditional software tools. Examples of unstructured data include "big data," social media posts, emails, customer reviews, and news articles. Unstructured data requires specialized tools and techniques.

[3] This included analysis of the DEA's ARCOS database, company internal transaction databases detailing shipments, and drug dispensing record databases.

[4] Inclusive of data produced from Apache Hive, No-SQL, Oracle, and other SQL databases.

10. I was retained in this MDL litigation to assist with ESI discovery discussions on March 18, 2024.

11. I signed the Protective Order in this MDL on July 23, 2024.

12. I have access to the produced documents, deposition transcripts and exhibits, and filings in this litigation.

13. I understand that on July 9, 2024, the Court ordered Uber[5] to produce certain data related to incidents of alleged sexual assault and sexual misconduct occurring in the United States from 2017 to 2022 which had been reported to Uber (Dkt. Nos. 683, 684) ("Incident Data").[6]

14. Relatedly, I understand that Plaintiffs have requested certain data from Uber's Safety Risk Assessment Dispatch ("S-RAD") program.[7]

15. I have been asked by counsel for Plaintiffs to opine on Plaintiffs' requests for S-RAD related data and Uber's response.

Uber's S-RAD Program

16. In 2017, Uber internally released early results of the program it called Safety Risk Assessed Dispatch or S-RAD.[8]

---

[5] Defendant Uber Technologies, Inc. (herein "Uber")
[6] I have directly accessed and queried the produced Incident Data.
[7] I have participated in multiple meet and confers ("M&Cs") between Plaintiffs' counsel and Uber's counsel regarding the production of both the Incident Data and S-RAD data.
[8] *See* UBER_JCCP_MDL_003385297

2025.09.22 DECLARATION OF JONATHAN JAFFE   3
IN RE: UBER TECHNOLOGIES, INC., PASSENGER SEXUAL ASSAULT LITIGATION, MDL No. 3084

[REDACTED]

21. The algorithm used to calculate the S-RAD Score has changed over time. Uber refers to each version of its S-RAD algorithm as a unique "S-RAD Model."

22. "Beta"[15] S-RAD Models refer to S-RAD Models in experimental testing stages before they are applied regularly to Uber trips.

---

[10] Uber refers to these potential matches of rider-driver as "Supply Plans."

[11] *See* UBER_JCCP_MDL_003719029 (S-RAD launched to 196 US Cities)

[12] *See* UBER_JCCP_MDL_004929682 column AK for sample S-RAD Scores.

[13] as opposed to a human-developed and coded; machine learning models tweak variables to find mathematical functions that most closely resemble multi-dimensional mathematical curves through an optimization process using a predefined set of training data (parameters with an expected output)

[14] There are dozens of parameters used in each S-RAD algorithm. *See* UBER_JCCP_MDL_003306684 and UBER_JCCP_MDL_005784097.

[15] beta is a software term that refers to a piece of software that has not been generally released; released software is commonly said to be "generally available" *see* https://www.ibm.com/think/topics/sdlc#:~:text=Developers%20might%20deploy%20software%20in,reaches%20general%20availability%20(GA) and https://support.google.com/a/answer/11202276?hl=en&src=supportwidget0&authuser=0

23. There have been one US[16] regional and two Global[17] S-RAD Models.

[black redaction blocks]

---

[16] *See* UBER_JCCP_MDL_005784097 and UBER_JCCP_MDL_005784097 referencing the US Production Model.

[17] *See* UBER_JCCP_MDL_002058117 and UBER_JCCP_MDL_005025910 (Global Model 1.0) and UBER_JCCP_MDL_003212988 and UBER_JCCP_MDL_003212988 (Global Model 2.0 released in 2024).

[black redaction blocks]

[21] *See* UBER_JCCP_MDL_003602459

[black redaction blocks]

[24] June 25, 2025 30b6 Deposition of Sunny Wong at 240:10-21.

How Uber Stores its S-RAD Data

28. During M&Cs in which I participated regarding S-RAD, Uber disclosed that its S-RAD data is stored in an Apache Hive[28] database as Apache Parquet[29] formatted files.

29. A "table" in an Apache Hive database is defined by a folder containing data files.

30. Uber disclosed that the S-RAD data is partitioned by date,[30] and searchable by the trip identifier.[31]

31. There are nearly 100 tables with S-RAD data.[32] It is not uncommon for Apache Hive databases to have thousands or tens of thousands of tables. Since each

---

[25] *See* UBER_JCCP_MDL_002059019

[26] Uber's definition of a city does not necessarily align with a governmental jurisdiction.

[27] As per Uber's counsel representation during our M&Cs, Uber sets two thresholds per city approximately once per 30 days; however, the time period in between settings may vary.

[28] https://hive.apache.org; Apache Hive is built on top of Apache Hadoop (*see* https://hive.apache.org/docs/latest/introduction-to-apache-hive/) to enable "easy access to data via SQL," which is a Structured Query Language

[29] Parquet is a data format that allows for extremely fast and efficient querying of millions and billions of records by columns.

[30] It is common practice to partition (divide into subfolders) Apache Hive tables by date to enable efficient querying by date range. In this situation, the date partitions allow queries to examine only the data relevant to the desired dates.

[31] The trip identifier is also known as the trip_uuid.

[32] *See* UBER_JCCP_MDL_003757150, UBER_JCCP_MDL_005320179, UBER_JCCP_MDL_003601911, and UBER_JCCP_MDL_000366010

"table" is self-contained, Plaintiffs' requests require examination of a small number of columns in a small number of readily identifiable folders.

32. Even though Defendants have asserted that a "30 day snapshot of just one of the items requested by Plaintiffs takes up more than 84 terabytes," this measurement includes all of the fields stored, not the limited fields requested. The fields Plaintiffs have requested are small in number and size. Apache Parquet files are columnar storage, that is, queries only scan and retrieve the bytes from the fields that are included in the query. As such, queries across even petabytes of data do not retrieve all of the data, but a fraction thereof. Moreover, queries are distributed such that multiple virtual compute processes divide up the work and execute simultaneously without much special configuration or effort.

S-RAD Data Produced to Date

33. Uber has produced the S-RAD Scores for each bellwether trip ("Bellwether Trip's S-RAD Score").

34. Uber has produced the parameters and values Uber used to compute the Bellwether S-RAD Scores ("Bellwether Trip S-RAD Parameters").

35. Uber has produced whether the Bellwether Trip was flagged and/or actioned by S-RAD ("Bellwether Trip S-RAD Status").

36. Uber has produced the S-RAD Trigger Rates over the 7 days preceding the Bellwether Trips only for the specific city in which the Bellwether Trip occurred ("Bellwether Trip 7-Day Preceding City Trigger Rates").

37. Uber has produced the average S-RAD score across other trips over the 7 days preceding the Bellwether Trips only for the specific city in which the Bellwether Trip occurred ("Bellwether Trip 7-Day Preceding City Average S-RAD Score").

38. Finally, Uber has produced the median 1-star rating of other drivers over the 7 days preceding the Bellwether Trips only for the specific city in which the Bellwether Trip occurred ("Bellwether Trip 7-Day Preceding City Driver Median 1-Star Ratings").

S-RAD Data Requested

Beta S-RAD Model Features

39. Plaintiffs have requested the parameters and relative weighting those parameters for each of the Beta S-RAD Models.

40. Without the parameters and relative weights for each Beta S-RAD Model, Plaintiffs cannot know what parameters Uber was testing for models that were not generally released. In my experience, extensive testing around a particular parameter is indicative of concerns about that particular parameter.

S-RAD Scores for all Trips in Incident Data

41. Since S-RAD was under development through much of the period covered by the Incident Data, I expect that out of the roughly 400 thousand trips in the Incident Data, less than 25% will have calculated S-RAD Scores.

42. Every trip in the S-RAD data has a date and unique trip identifier.[33]

---

[33] Trip identifiers are globally unique.

43. I have queried similar Apache Hive databases with hundreds of billions of records similarly partitioned by date for fields to retrieve fields by records identifiable with a unique identifier.

44. Given the volume of data, it is my opinion that using standard querying methods (Presto[34] or Apache Spark[35]) it would take no more than 30 minutes[36] to pull the S-RAD Scores for each of the expected 100,000 or less unique trips in the Incident Data where Uber had calculated an S-RAD Score.

45. Without the S-RAD Scores across all Trips in the Incident Data, Plaintiffs cannot construct distributions of S-RAD Scores for similar events both before and after, nor compare S-RAD Scores for more and less serious events before and after in the same city or similar cities or dissimilar cities.[37]

S-RAD Thresholds

46. As described *supra* in ¶ 27, S-RAD Thresholds (daytime and nighttime) are set once every 30 days for each S-RAD Model for each city as defined by Uber.

47. Given that there are less than 400 cities in the Incident Data,[38] given that S-RAD Thresholds are set on average twelve times a year, given that there are on average two thresholds per city (daytime and nighttime), and given that the S-RAD program has been operational less than 10 years, the total number of S-RAD Thresholds would fit on a single five-column spreadsheet.[39]

---

[34] https://prestodb.io/docs/current/overview.html
[35] https://spark.apache.org/docs/latest/api/python/index.html; also commonly referred to as PySpark
[36] and more likely than not under 2 minutes given the volume of data
[37] For example, were S-RAD Scores lower in larger or smaller cities with more serious events.
[38] Based upon city identifiers in the Incident Data.
[39] 400 * 12 * 2 * 10 = less than 96,000 potential historical thresholds

48. Given the volume of data, it is my opinion that using standard querying methods it would take no more than a few minutes to pull the S-RAD Thresholds historically for all the cities in the Incident Data.

49. Without historical S-RAD Thresholds, Plaintiffs are unable to know the distribution of thresholds overall, compare thresholds in similar cities to the Bellwether Trips, to know whether the threshold was historically lower than before the Bellwether Trips or whether and how it changed after the Bellwether Trips.

S-RAD Trigger Rates

50. As described *supra* in ¶ 24 and disclosed in M&Cs in which I participated, S-RAD Trigger Rates are computed after-the-fact percentages describing the percentage of trips where a trip was flagged and/or actioned by S-RAD.

51. Uber computed these values for the Bellwether Trips but did not provide how it computed these values.

S-RAD Slip-Throughs

53. I understand that Plaintiffs have requested both Type 1 S-RAD Slip Throughs and Type 2 S-RAD Slip Throughs for the trips in the Incident Data.

---

[40] *See* UBER_JCCP_MDL_003224079 and UBER_JCCP_MDL_003503729
[41] *See* UBER_JCCP_MDL_002059019, UBER JCCP MDL 002340895, and UBER JCCP MDL 002658347

54. As discussed *supra* in ¶ 41, it is my opinion that the volume of data at issue is minimal.

55. Furthermore, it is my opinion that pulling all ride checks[42] and nudges[43] for the trips in the Incident Data would take under 30 minutes and fit in a single CSV[44] file.
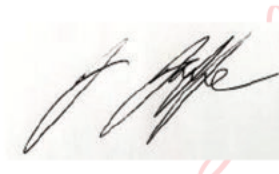
S-RAD Supply Plans

56. I understand from what Uber disclosed in M&Cs in which I participated that S-RAD Supply Plans are only available for the past 30 days.

57. Supply Plans are necessary to determine what alternatives were available especially when a Trip was flagged or actioned by S-RAD.

I declare under penalty of perjury that the foregoing is true and correct.

Digitally signed by Jonathan Jaffe
Date: 2025.09.22 15:23:57 -04'00'
Adobe Acrobat version: 2025.001.20643

Executed on September 22, 2025.

_____
Jonathan Jaffe

---

[42] A "ride check" is a prompt to a rider and/or driver during a trip from the Uber application inquiring whether the rider is in need of assistance. Ride checks require the recipient to respond.

[43] A "nudge" is a message to a rider and/or driver during a trip in the Uber application intending the recipient to take some action. Nudges do not require the recipient to respond. Nudges are also referred to as push notifications. There are S-RAD specific nudges that according to Uber during our M&Cs were not deployed to the US; however, I understand Plaintiffs' request to be broader than that to all nudges sent during Trips in the Incident Data.

[44] comma separated value, a human-readable text format